

A tartalom szerinti információkeresés az interneten

I. Indexelőszolgáltatások

Az internetes keresőszolgáltatások mind rugalmasságban, felhasználóbarát felületek dolgában, mind az információs kínálatban messze fölülmúlják a távolsági online szolgáltatások adta lehetőségeket. Mindez kihívás az információkeresés és osztályozás számára, amely az internet megjelenésével történetének legjelentősebb fejlődése előtt áll. A keresőszolgáltatásokat kezdettől fogva ugyanaz a kettősség jellemzi, mint minden hagyományos tartalom szerinti kereső és rendező rendszert: kialakultak a természetes nyelven működő, olykor már szabványosított szótárakat (tezauruszokat) is alkalmazó indexelőszolgáltatások, és a hierarchikus osztályozási rendszereket alkalmazó internetkatalógusok. Frissen kialakult szóhasználatukat megkíséreljük összehangolni a dokumentációskönyvtári terminológiával. Az első részben az indexelőszolgáltatásokkal, a második részben az internetkatalógusokkal foglalkozunk, végül röviden kitérünk az elsődleges és másodlagos elektronikus dokumentumok formátumaira is.

1. Információkeresés az interneten, avagy világméretű hozzáférés a tömegek számára

Az adatbázisokkal, akárcsak a hagyományos könyvtári katalógusokkal nemcsak a szakembereknek, hanem az alkalmi felhasználóknak is boldogulniuk kell. Az online információkereső rendszerben a felhasználónak nem kell ismernie magát a keresőrendszert, ahogy a személyautó vezetőjének sem kell értenie járműve szerkezetéhez. Olyan felhasználói felületet kell létrehozni, amelynek szemiotikai struktúrája az adott kultúra megszokott, mindennapi jelrendszerének felel meg, azaz a természetes, laikus gondolkodásnak. A felhasználónak erre a „végfelületre” (end user interface) van szüksége, és ha a szolgáltatók ezt biztosítják, akkor a könyvtártudomány negyedik rangban van törvényének tesznek eleget: „*Kíméljük az olvasó idejét!*”

Az ide vezető fejlődés már közvetlenül a háború után elkezdődött, és a hipertext (hypertext) feltalálásán keresztül vezetett el az interneten megvalósult globális információkereséshez. (A fejlődést kizárólag a tartalom szerinti – például jelentéssel bíró szavak alapján végzett – információkeresés szemszögéből tárgyaljuk, ezért a hálói

zati rendszerekkel és az internet egyéb vonatkozásaival nem foglalkozunk.)¹

Vannevar Bush, aki a háború alatt az amerikai tudósok „hadseregét” irányította, és ezzel a győzelem egyik fontos, bár jobbra ismeretlen alakjává vált, 1945-ben fogalmazta meg először, hogy az információkeresés folyamatának (akkor még ezt a kifejezést – information retrieval – nem használták) az asszociatív kapcsolatokon kell alapulnia. Az *Úgy, ahogy gondolkodhatunk* [1] és az *Endless horizons* [2] (*Végtelen láthatárok*), majd a húsz év múlva újrafogalmazott *Memex revised* (*Módosított Memex*) [2] című tanulmányaiban ő használta először az összekapcsolt szövegblokkok fogalmát, ő vezette be a „link” (ebben az esetben a releváns szöveghelyekre utaló egyszerű kapcsolatjelölő: csatoló, kapocs, utaló, mutató, hivatkozás, ugrópont) és a nyomvonal, valamint a háló kifejezéseket a textualitás új elképzelésének a leírására. Konceptiója a gépesített, határtalan kapacitású, mindenféle dokumentumokat tartalmazó iratgyűjtemény és könyvtár, amely a felhasználó számára gyors, asszociatív keresést tesz lehetővé. A fél évszázada megálmodott elektronikus, hálózati könyvtár feltételei napjainkra értek meg.

Munkássága nagy hatással volt Douglas Engelbartra (az egér és az ablaktechnika feltalá-

lójára) és a hipertext későbbi úttörőire, mint Theodor Holm Nelsonra, és a Brown University Információ- és Tudománykutató Intézetének (Institute for Research of Information and Science = IRIS) kutatócsoportjára, az Intermedia megalkotóira. Nelson eszméjének lényege, hogy a lineáris szövegfolyamon belül kisebb szövegrészeket kapcsolt össze. Ezek a kapcsolatok a szöveget keresztül-kasul behálózták, az olvasó maga határozhatta meg, milyen legyen az általa tanulmányozott szöveg szerkezete. Megszületett a nemlineáris szöveg eszméje. George P. Landow, a Brown University angol irodalom és művészettörténet professzora a hipertext és az internet keletkezéstörténetével foglalkozó művében a következőképpen világítja meg a hipertextes, „központ nélküli” technikának a gyökereit.²

„Amikor az olvasók szövegek hálójában vagy hálózataiban haladnak előre, folyamatosan változtatják kutatásuk vagy tapasztalatuk középpontját – s ezáltal a fókuszot vagy a szervező elvet is. Más szóval, a hipertext olyan korlátlanul újra középpontosítható rendszerként szolgál, melynek ideiglenes fókuszpontját az olvasó jelöli ki, akiből ennek ellenére más értelemben válik valódi aktív olvasó. A hipertext egyik alapvonása, hogy egymással összekapcsolt (Roland Barthes által lexiáknak nevezett) szövegtestekből áll, melyek nem egyetlen fő szervező tengely mentén kapcsolódnak. Más szóval, a metaszövegnek vagy dokumentumsornak – annak a dolognak, entitásnak, amely a nyomdai technikában meghatározza a könyvet, a művet vagy a szöveget – nincs középpontja. Igaz ugyan, hogy a középpont hiánya problémát okozhat az íróknak és az olvasónak is, ám a hipertextet használva mindenki saját érdeklődését teszi meg kutatása pillanatnyi de facto szervező elvének (vagy középpontjának). A hipertextet olyan rendszerként tapasztaljuk meg, mely korlátlanul középpont nélkülivé tehető és újra középpontosítható részben azért, mert a hipertext átmeneti középponttá, a tájékozódást és a továbbhaladást segítő könyvtári katalógussá alakít bármely dokumentumot, mely egynél több kapcsolódással – csatolóval (linkkel) – rendelkezik.

A nyugati kultúra jóval a számítástechnika előtt ismerte már a hálózatra kapcsolt valóság félig-meddig mágikus kapuit. A bibliai tipológia, mely olyan fontos szerepet játszott az angol kultúrában a XVII. századtól a XIX. századig, a krisztusi elrendelés típusainak és előjeleinek kategóriáiban gondolta el a bibliai történetet. Vagyis Mózes, aki a saját jogán létezett, létezett Krisztusként is, aki beteljesítette a próféta jóváindulását. Számítlan XVII. századi és viktoriánus prédikáció, traktátus és szövegmagyarázat demonstrálja, hogy bármely személy, esemény vagy jelenség mágikus ablakként szolgált az emberi üdvözülés isteni rendjének összetett szemiotikájában. A jelentős eseményeket és jelenségeket egyidejűleg több valóságban vagy valóság-szinten megjelenítő bibliai típushoz hasonlóan az egyes lexiák is szükségszerűen utat nyitnak a kapcsolatok hálózataiba. Feltéve, hogy az evangélikus protestan-

tizmus Amerikában megőrzi és továbbfejleszti a bibliai szövegmagyarázatnak ezt a hagyományát, cseppet sem meglepő, hogy a hipertext első alkalmazásai között ott volt a Biblia és az exegetikai tradíció.

Valamennyi hipertextrendszer lehetővé teszi, hogy az olvasó maga válassza ki a kutatás vagy a tapasztalat középpontját. A gyakorlatban ez az elv azt jelenti, hogy az olvasó nincs bezárva semmiféle szerkezetbe vagy hierarchiába [15].³

T. H. Nelson 1965-ben írta le a hipertext nevet, és határozta meg – a felhasználó szemszögéből – a fogalmát:

„Írott vagy képi anyagok olyan komplex összeköttetése, amit papíron nem lehet kényelmesen megalkotni. Összefoglalókat és térképeket tartalmazhat a benne szereplő anyagokról és ezek egymáshoz való viszonyáról; tartalmazhatja az anyaggal foglalkozó tudósok megjegyzéseit és lábjegyzeteit is.”⁴

Nelson egyben elkezdte a Xanadu nevű, máig meg nem valósult, maximalista hálózati hipertext-rendszerének és az általa Egységes Adatstruktúrának (Unified Data Structure) nevezett formátumnak a tervezését is. Ahogy V. Bush a felhasználóbarát számítógépes végfelületek, az ablak-technika és a hipertext feltalálójának körét, úgy Nelson Xanadu terve és Egységes Adatstruktúrája programozók kis, de lelkes körét befolyásolta tartósan. (Áttételesen még a bibliográfiái adatcsereformátum létrehozását is inspirálhatta.)⁵

A Xanadu név Coleridge egyik költeményéből származik: az „irodalmi emlékezet mágikus helyét” jelöli, ahol minden megőrződik. Ahogy Sugár János fogalmazott: „A névválasztás is jelzi a hipertext eredendő és mély irodalmi gyökereit. Talán a hipertext az első civilizációs vágyálom, melyet a rohamosan fejlődő számítástechnika valósít meg.”⁶ Figyelemre méltó, hogy csak akkor kerülhetett sor a hipertext Nelson által elképzelt alkalmazására, amikor az internet hálózata kialakult.

Miközben Nelson a Xanadu megalomániás tervét kergette, Andries van Dam 1967–1968 között elsőként ténylegesen működő hipertextrendszert készített. 1969-ben pedig az Egyesült Államok hadügyminisztériumának rendelkezésére megszületett az ARPANET (Advanced Research Project Agency Network), az internet őse. Annak érdekében hozták létre, hogy atomcsapás esetén se szakadjon meg a kommunikáció az amerikai kormány- és katonai szervek között. Az internet elve egyszerű: a hálózatra nincs központja, részei egymástól függetlenül működhetnek, mivel minden csomópont egyenrangú. A megcímezett adatcsomagok útja teljesen közömbös, csak az eredmény számít: a csomagok csomópontokról csomópontokra vándorolnak, míg el nem érik a címzettet. Hiába semmisült volna meg számtalan csomó-

pont, a küldemények a háló megmaradt csomópontjait érintve járhatták az útjukat. Mivel idővel egyre több nem katonai intézmény is csatlakozott a hálózathoz, a nyolcvanas évek elején katonai részét különválasztották, és ami megmaradt, ahhoz teljesen szabaddá tették a csatlakozást. Az internet diadalútja elkezdődött.

A nyolcvanas évek végére már csak a könnyen kezelhető, grafikus kezelőfelület hiányzott, olyan felhasználóbarát „műszerfal”, amely a legostobább végfelhasználó számára is lehetővé teszi a keresést az internetben. Ekkor jelent meg a színen Tim Berners-Lee, aki 1989-ben az Európai Részecskefizikai Laboratóriumnak (CERN) – saját bevallása szerint – a Xanadu inspirációjára javasolta a World Wide Web tervét. (A Xanadu programot viszont, mint annyi más úttörő, számos irreális vonást tartalmazó kezdeményezést, az AutoDesk Company. 1992-ben – miután közel 5 millió dollárt költöttek rá – feladta.⁷) Akárcsak Nelson, Berners-Lee sem gazdagodott meg találmányából, mivel szabadalmi és copyrightigényeinek bejelentését mellőzve, eredetileg csak a tudományos közösség számára akart olyan eszközt létrehozni, amellyel a hipertextes közleményeket korszerű szerkezetben lehetett megjeleníteni és olvasni az interneten.

Az internet jelentősége, hogy távolsági online hozzáférést tesz lehetővé a laikus „tömegek” számára. Ebből nem következik, hogy nincs már szükség a speciális szakterületekre vonatkozó, elsősorban pénzért szolgáltatató online adatbázisokban végzett kereséskor a részletesebb információkereső szakmai ismeretekre. Mint minden fejlődésben, itt is differenciálódás játszódik le: az internettel a használat újabb szintje jelent meg anélkül, hogy a korábban kialakult használati módok érvényüket vesztenék. Ahogy nem szűnik meg a nyomtatott dokumentumok használata sem az elektronikus dokumentumok megjelenésével (ennek ellenkezőjét legfeljebb az internet terjedésében érdekelt nyomásgyakorló csoportok tagjai terjesztik).

Mivel az internet több szakterület (adatátvitel, programozás, katalogizálás, osztályozás, információkeresés) metszéspontjában fekszik, melyeknek mind önálló szakmai nyelvük van, az internettel átfogóan foglalkozó szakirodalomban az egyes szakterületek terminológiáját olykor felszínesen vagy önkényesen használják. A könyvtártudományban és dokumentalisztikában, később meg az online információkeresésben történetileg kialakult terminológia értelemszerűen fölhasználható az internetben végzett tartalmi feltárára és keresésre, de ez csak lassan valósul meg.⁸

1.1 Az internet méretei

Barabási Albert László, az indianai Notre Dame Egyetem tanára kiszámította, hogy két, a jelenleg 800 millió weboldal közül kiválasztott tetszés szerinti HTML-dokumentum legfeljebb 19 hivatkozási ugrásnyira van egymástól. Bármit keresünk is a csatolók („linkek”) segítségével szörfölve, az átlagosan nincs messzebb, mint 19 csatolóról csatolóra megtett lépés – legalábbis statisztikai szempontból. A számítógépek, amelyek az internetet fenntartják, olyan szorosan összekapcsolódtak már, hogy a 800 millió dokumentum közül még a legtávolabbi is rövid idő alatt elérhető.

„A bonyolult topológiájú véletlen hálózatok általánosan elterjedtek a természetben, és segítségükkel olyan eltérő rendszerek modellezhetők egyformán, mint a World Wide Web vagy a társadalmi és gazdasági rendszerek. Újabban az is kiderült, hogy a legtöbb rendkívüli nagyságú hálózat topológiai információi skálafüggetlen jellemzőkkel írhatók le. Megvizsgáltuk ezeknek az újabban ismertté vált skálafüggetlen modelleknek a skálátulajdonságait, melyek a kisenergiájú eloszlásokon alapuló összefüggőségét (konnektivitását) megmagyarázhatják. A jelentésmező elméletet alkalmazva előre jelezhetjük a gráf csúcsainak növekedési dinamikáját, és kiszámíthattuk a jelentésmező elmélet alapján az összekapcsoltság eloszlását és a mértékfüggvényt. Az eredményeket a weben próbáltuk ki. [...]”

Olyan robotprogramot készítettünk, amely lényegében a web egy részét feltérképezte. Első lépésben adatbázisba gyűjtötte az egyes honlapok csatolóit, majd követi azokat a hivatkozott honlapokon, és ezt folytatja. A kapott adatokat statisztikai módszerekkel értékeltük: meghatároztuk annak valószínűségét, hogy a vizsgált dokumentumra megadott számú honlapról hivatkoznak, illetve e honlap ugyanannyi oldalra hivatkozik. A véletlen hálózatokra jellemző valószínűségi eloszlást vártunk. Ez azt jelentette volna, hogy a legtöbb honlapon mondjuk 10–20 csatoló lesz majd más weboldalakra. Mi voltunk a legjobban meglepve, amikor ehelyett egészen mást találtunk, azt, hogy a csatolók számának eloszlása hatványfüggvény, ami önszervező rendszerekre jellemző, és arra utal, hogy sok honlap van, amelyről több ezer csatoló indul ki, és ugyanakkor rengeteg olyan honlap van, amelyre hihetetlenül sok más honlap mutat. Noha bármely weboldal-tulajdonos teljesen szabadon döntheti el, hány csatolót helyez el a honlapján, a teljes hálózat mégis általános törvényszerűségnek engedelmeskedik. [...]”

Az eredmények a keresőrendszerek tervezésében hasznosíthatók. A mai keresőprogramok helyett idővel talán kidolgozhatnak az új felfedezésen alapuló, intelligensebb keresési módszert, amennyiben kihasználják a háló összefüggőségét, és a felhasználó által kívánt információt az azonosított, legfeljebb tizenkilenc csatolót végigkövetve keresik meg.⁹

A vizsgálatokra használt program tehát adott HTML-dokumentum összes csatolóiból kiindulva

addig követte az adódó csatolókat, ameddig csak újabbat talált. A folyamatot rendkívül sokszor megismételték, majd statisztikai módszerekkel kiszámították két HTML-dokumentum között az átlagos csatolórészek számát. Ez a háló két pontja közötti átlagosan legrövidebb „távolság”, melyet a háló „átmérőjének” neveztek el. 800 millió HTML-dokumentum esetén eszerint két véletlenszerűen kiválasztott pont közötti átlagos távolság 19 csatolórész.

Ha a háló jelenlegi méreteinek tízezerszeresére növekszik, akkor a hatványtörvény megmondja, hogy legfeljebb huszonegy csatolórészre lesz szükségünk egy honlapról tetszőleges másik honlap elérésére.

Az eredmény egyben megvilágítja a web relatív kommunikációs korlátait. Az emberiség ugyancsak önszervező rendszernek tekinthető, melyre vonatkozóan az összekapcsoltság kiszámítható. Barabási utal rá, hogy a ma élő 6 milliárd emberre vonatkozóan ezek a számítások kimutatták: legfeljebb hat lépésben minden embernek más embereken keresztül kapcsolata van bárkivel a világon. Azaz mindenki ismer olyan embert, aki megint csak ismer olyan embert stb., és a sorban a hetedikhez jutva az emberiség bármelyik tagjával áttételesen kapcsolatban vagyunk. A web „kompaktsága” tehát az emberiségéhez képest alig harmadannyi.

2. A keresőrendszerek története

Az internet forrásainak eléréséhez kezdetben meglehetősen körülményes, a laikus felhasználó számára nehezen vagy alig használható eszközöket alkalmaztak. Csak arra voltak jók, hogy a kapcsolat lehetőségére a szabványos feltételeket megteremtsék, és elvégezzék az indexelést.

- Az *FTP (File Transfer Protocol)* az adatátvitel általános szabványa, s egyben program biztosítja, hogy a hálózat számítógépei között egységes formában jöjjön létre az állománycseré.
- A *WAIS (Wide Area Information Servers)*, az *Archie*, illetve a megfelelő későbbi programok (pl. *Apache*, *INQUERY*) hozták létre a másodlagos adatállományokat (indexeket, relevancia-adatokat és leírásokat), amelyek az internetre kerülő dokumentumokra utalnak. Ezek a szervertől indexelő rendszerek.

Mivel a háttérben működnek, démonoknak is nevezik őket. Használatuk bonyolult, és hiányzott mögülről az intézményes háttér. Az áttörés két olyan rendszer megszületéséhez fűződik, amelyek bizonyos szempontból homlokegyenest ellentétei egymásnak.

- 1991-ben először *Gophereket* készítettek. Ezt a szöveges, menüszerkezetű információs hálózatot leginkább a hierarchikus felépítésű tartalomjegyzékekhez lehetne hasonlítani. A Gopherben a felhasználó szöveges (és csak szöveges) állományokat nézhetett meg és tölthetett le. Legismertebb keresőeszköze a *Veronica* integrált menülekérdező és indexelő rendszer (Very Easy Rodent-Oriented Net-wide Index to Computerized Archives).

- Még ugyanebben az évben, Tim Berners-Lee kísérleteiből kiindulva megszületett a *World Wide Web* (web, WWW) hipermédia információs hálózat üzemeltető formája. Ebben a rendszerben a hipertext jóvoltából az információforrásokat hipertext csatolók (hyperlinks) formájában teljesen szabadon lehet egymással összekapcsolni. A kapcsolódó és megjeleníthető információforrások nemcsak szövegek, hanem képek, hangok és mozgóképek is lehetnek. A rendszeren belül nem érvényesül semmiféle hierarchikus rendező elv, minden forrás egyszerre több más forrással is összefügghet és fordítva (azaz a kapcsolódások szerkezete M:N). A weben belül hamarosan kialakultak különféle *keresőszolgáltatások*, mint például az *AltaVista*, *HotBot*, *Infosec*, *Magellan*, *Excite*, illetve *Yahoo!*, Magyarországon a *HUDIR*, illetve a *Heuréka* és az *AltaVizsla*. A web formájában végre megvalósult az általánosan hozzáférhető és az asszociatív gondolkodással összhangban álló felhasználói felület, amelyet 1945-ben Vannevar Bush megálmodott. Benne minden addigi egységesítő (FTP), feldolgozó és keresőeszközt (WAIS stb.) integráltak.

A versenyből a Gopherrel szemben néhány év alatt a web került ki győztesen, de a Gopher-WWW kettősség nem véletlen jelenség, hanem a rendező rendszerek kettős természetével függ össze (a dichotómiára a későbbiekben még visszatérünk).

Az internet fejlődését másik kettősség – a kereslet-kínálaté – is meghatározta. Ennek következtében alakult ki a háló, s ezzel a web két „oldala”: a tartalomszolgáltatás és a keresőszolgáltatás.

3. Tartalomszolgáltatás

Az interneten nem volna mit keresni, ha nem lenne *tartalomszolgáltatás*. Ezen egész általánosan az információ (az adatok) elhelyezését értjük az interneten, főleg hipertextes HTML-dokumentumok formájában.

A távoli hozzáférésű források egyre nagyobb része a http kommunikációs szabályai alapján elérhető dokumentum. Köztük ma szinte kizárólagos szerepe van a HTML-formátum szerinti szerkezetű dokumentumoknak, noha elképzelhető, hogy a jövőben a http-n belül más (pl. xml) szerkezetű dokumentumok is el fognak terjedni. A többi kommunikációs protokoll és szabály szerint elérhető távoli hozzáférésű forrásoknak (pl. Gopher-, telnet-, ftp-dokumentum, elektronikus levél) alig van jelentősége, és ezért a továbbiakban, ha az internet dokumentumairól van szó, csak HTML-dokumentumokról beszélünk.

A kisebb vagy nagyobb tartalmi egység kezdő- vagy belépőlapját honlapnak (home page, otthallap) nevezik. Lényegében mindenki – akarva-akaratlan – tartalomszolgáltató, akinek honlapja van az interneten (így a személyes honlapok tulajdonosa is); a keresőszolgáltatások programjai elkerülhetetlenül indexelik az elérhető HTML-dokumentumokat, nem téve különbséget a kereskedelmi céllal végzett tartalomszolgáltatás és az egyéb (pl. intézményi, személyi) honlapok között.

A gyakorlatban tartalomszolgáltatáson a kereskedelmi célú információelhelyezést értik, mellyel erre szakosodott cégek foglalkoznak. A tartalomszolgáltatásban fontos szerepet betöltő webhelyek honlapjára külön elnevezések születtek (honlap, otthallap, hálózsem). Ha a sokféle elérhetőség első-

sorban kereskedelmi, szolgáltató vagy legalábbis professzionális, intézményi jellegű, akkor portálról (portáldalról, portálszolgáltatásról) beszélnek.¹⁰

Léteznek tartalomszolgáltatók, akik a különféle keresőszolgáltatásokról tájékoztatnak, gyakran értékelve is ezek minőségét. Ezekből kiderül, hogy a keresőszolgáltatások száma 1999 végén több száz volt, és lehet, hogy előbb-utóbb megközelíti az ezret. Kisebbségi hányaduk ún. gyűjtőszolgáltatás, azaz az összegyűjtött, csoportosított keresőszolgáltatások közvetlenül kérdezhetők le, anélkül, hogy át kellene térni magára a lekérdezett szolgáltatásra. A portáldal példaként ilyen szolgáltatás szerepel az 1. ábrán.¹¹

A rendszerek működését fizikai berendezések a „helyek” (site). A webszervergép például hely, és eme igazi helyen belül afféle virtuális „részhelyek” a gépen található információforrások (pl. adatbázisok, hirdetőtáblák, honlapok). A „tér” a hálózat, a „helyek” pedig a kiindulási, fizikai szinten gépek, a további szinteken pedig a különféle „részhelyeket” képviselő webhelyek, ftp-helyek, hírcsoportok helyek stb., melyek az információforrások szerepét játszhatják. A hálózat gépeit, illetve a virtuális gépek „helyét” (doméneket), és azon belül az információforrásokat (a helyeket és „részhelyeket”) a „térben” szabványosított formában írt azonosító, az URL (Unified Resource Locator) jelöli. Egy szer-

BSZ
Bibliotheksservice-Zentrum Baden-Württemberg

Home
Datenbanken
Publikationen
Projekte
Bibliotheksdienste
Links
KEBweb

Suchdienste
Auskünfte
Verkehr
Wetter
Jobs & Soziales

Internet
Presse
Gesetze & Recht
Frauen

Links aus aller Welt
gesammelt vom Bibliotheksservice-Zentrum Baden-Württemberg (BSZ)
Attention please! We moved!
Click here for new URL: <http://www.bsz-bw.de/links/suchdienste.html>

Suchdienste im Internet
Search & Find in the Internet

NEW! What's new
[WWW-Suchdienste international | International Search Engines](#)
[WWW-Suchdienste deutschsprachige Seiten | Germany](#)
[WWW-Suchdienste europäische Länder | Europe](#)
[WWW Virtual Libraries](#)
[E-Mail-Adressen, Personen](#) * [Mailinglisten, Newsgroups](#)
[Suche nach WWW-Servern](#) * [Städte und Orte im WWW | Cities](#)
[Suche nach Bibliotheken, Katalogen, Bibliotheksinformationen | Libraries](#)
[Suchmaschinen für Dublin Core Metadaten | Metadata Search Engines](#)
[Suche nach Hochschulen, Instituten, Schulen | Universities, Education](#)
[Lieferantenverzeichnisse](#)
[FTP-Suchdienste](#)
[Link-Börsen und Homepage-Anmeldedienste](#)
[Meta-Indices, Suchdienste-Verzeichnisse, Einführungen | Directories](#)

Nachschlagewerke, Lexika, Dictionaries

1. ábra Keresőszolgáltatásokról tájékoztató, kétnyelvű tartalomszolgáltató portálja

vergépi internethelyen általában nagyon sok rész-hely (tartalomszolgáltató, honlap, azaz webhely) található.

4. Keresőszolgáltatás

4.1 Általános jellemzők

4.1.1 Meghatározás

Az elsődleges adatokat tartalmazó dokumentumokat csak a részben belőlük nyert (pl. cím, kiadó), részben intellektuálisan megállapított (pl. besorolási adatok, deskriptorok, jelzetek) másodlagos adatok alapján lehet keresetileg tárolni. A másodlagos adatok egy-egy dokumentumra vonatkozó összessége alkotja a dokumentumleírást vagy dokumentumrekordot, könyvtári szabványoknak megfelelő formájuk a bibliográfiai tétel. Az interneten is meg kell különböztetnünk az elsődleges dokumentumok szerepét játszó HTML-dokumentumokat, és a rájuk vonatkozó másodlagos vagy metaadatokból felépülő HTML-rekordokat. HTML-rekordon tehát a HTML-dokumentumról készült, az oldal másodlagos adatait tartalmazó információtelet értjük. Ez utóbbiakat az internetes keresőszolgáltatások hozzák létre annak érdekében, hogy a felhasználók keresni tudjanak.

A HTML-dokumentumokhoz való hozzáférést biztosító keresőszolgáltatások jelentős része a hirdetésekben tartja fenn magát, és nagy részük ingyenes, kisebb részük használatáért (pl. a NetFirst) fizetni kell. A hirdetések a szolgáltató lapjain megjelenő csatolókon keresztül érhetők el; ezeket kiválasztva jut el az érdeklődő a hirdetés feladójának HTML-dokumentumára. Minél többen használják az adott keresőszolgáltatást, minél többen keresnek a segítségével, annál jobban vonzza a hirdetőköt, annál több a jól fizető hirdetés. A szolgáltatások fejlődését ma elsősorban ez határozza meg. A keresőszolgáltatások érdekelték abban, hogy a felhasználók könnyebben és eredményesebben kereshessenek, ezért a választék növelése érdekében óriási adatbázisok alakultak ki. Így érik el, hogy sokan használják őket, ami kihat a hirdetések számára. Mindez emlékeztet a sajtó világára.

Tágabb értelemben keresőszolgáltatások a webinterfészrel rendelkező online nyilvános adatbázisok is. Ezek elsődleges dokumentumai nem HTML-formátumúak, a szolgáltatáshoz dinamikusan lefordítják őket erre a formátumra. Általuk valójában távolsági online információszolgáltatás valósul meg az interneten (a nagy online szolgáltatók webopciót biztosítanak a használatukhoz, mint amilyen például a DIALOG Web). Az adatbá-

zisok ezáltal a klasszikus (telnetes) adatátviteli hozzáférés mellett, illetve helyett a weben keresztül is elérhetők. E szolgáltatások az osztályozás és az információkeresés szempontjából változatlanul „hagyományosak”, és többnyire nem ingyenesek. Túlnyomórészt ellenőrzött információkereső-nyelvi szótárt (tárgyszójegyzéket vagy tezauruszt, illetve osztályozási rendszereket) használnak bennük, az információk dokumentációs egységek (dokumentumok leírásai), a tartalmi feltárást intellektuálisan végzik, és az alkalmazott dokumentációs adatbázis-kezelő rendszerek jóvoltából a keresési lehetőségek sokkal fejlettebbek, mint az internetes kereskedelmi keresőszolgáltatások kizárólag indexekre vagy katalógusokra alapozott keresési lehetőségei.

Ugyancsak tágabb értelemben keresőszolgáltatásnak tekinthetők a speciális forrásokat egyetlen adatbázisból szolgáltató rendszerek, mint a webtelefonkönyvek, elektronikus menetrendek, elektronikus postai és egyéb címtárak stb., elektronikus szótárak, hirdetések, üzleti információk, humorlapok stb. Ezek sem HTML-dokumentumokat szolgáltatnak, hanem tényadatokat (faktografikus információkat). Jelentős részüket a globális keresőszolgáltatások segítségével is le lehet kérdezni. (Részletesebben a „Speciális adatbázisok” című fejezetben – 2. rész – tárgyaljuk őket.)

A továbbiakban csak a HTML-dokumentumokhoz való hozzáférést biztosító keresőszolgáltatásokkal foglalkozunk. Ezekben az osztályozás és információkeresés szempontjából jelentős új fejlemények figyelhetők meg.

A gyűjtőkör szempontjából a szolgáltatások többsége *globális*, azaz – legalábbis elvileg – az egész háló a gyűjtőköre (pl. AltaVista, Yahoo!), kisebbik része *nemzeti (állami)*, *regionális* vagy *meghatározott nyelvre* korlátozza a gyűjtőkörét (pl. AltaVizsla, HUDIR). Egy részük *egyetemes*, azaz mindenfajta és mindenféle tartalmú HTML-dokumentum a gyűjtőkörükbe tartozik, másik részük csak *speciális* tartalmú (pl. a WWW Women csak női tárgykörű) vagy speciális típusú (pl. a MusicSearch csak zenei) HTML-rekordokat szolgáltat. Az utóbbihoz tartoznak a szakterületi információs kapuszolgálatok (subject based information gateways) is.

Egyes szolgáltatások a gyorsaságukkal, mások a keresési eszközök gazdagságával, megint mások a feldolgozott állományuk nagyságával tűnnek ki. Vannak a relevancia szempontjából megbízhatóbbak, és kevésbé megbízhatóak. Mindez az erős versenyben állandóan változik. Gyakran jelennek meg elemzések a hálón, amelyekből tájékozódni lehet az aktuális helyzetről, de még nem alakultak ki megbízható tudományos módszerek az értékelésre (a hatvanas évek elején a hagyo-

mányos információkeresés hatékonyságára vonatkozó cranfieldi vizsgálatokhoz [10] hasonló jelentőségű elemzések még váratnak magukra).

4.1.2 Milyen keresőszolgáltatások léteznek?

Bármelyik nagyobb keresőszolgáltatásnak feltehetjük a kérdést, hogy hol található értékelés a keresőgépekről? Az AltaVistának például az „evaluation of search engines” vagy „Bewertung von Suchmaschinen” láncot, illetve összetett kereséshez a („search engines”) AND evaluation vagy Suchmaschinen AND Bewertung keresőkérdést, és válogathatunk az információk között. A szolgáltató rendszerek minőségéről mindig akadnak naprakész vizsgálatok, melyeket az interneten publikálnak (pl. [18]).

A szolgáltatások közötti nagy különbségek miatt nem szerencsés rangsorolni a teszteredményeket. A különféle keresési célokra különféle induló szolgáltatások vehetők igénybe. A MetaCrawler működtetői által végrehajtott vizsgálat azt jelezte, hogy pillanatnyilag egyetlen nagyobb keresőszolgáltatás sem képes a források 45%-ánál többet feltárni. Alig akad tehát olyan kérdés, amelyre egyszerre több szolgáltatással végzett keresés nélkül érdemben válasz kapható [13].

4.1.3 A rendező rendszerek kettőssége az interneten

Az információkeresés és -osztályozás szempontjából a keresőszolgáltatások két fő típusa alakult ki: az indexelő (vagy „keresőgépes”) szolgáltatások, és a saját adatbázist kezelő, a piacon az előbbinél valamivel korábban megjelent internetkatalógusok (vagy böngészőszolgáltatások). Számos szolgáltatásban egyre inkább mindegyik típussal találkozhatunk.

A két rendszer jóformán egyidejű megszületésében és ellentétében sajátos, az osztályozási rendszerek (és egyben az információkereső nyelvek) korai történetére emlékeztető logika köszön vissza, amely időközben a Gopher és a web kialakulásával kapcsolatban is megfigyelhető volt. Az indexelőszolgáltatások az analitikus (individualizáló, nem hierarchikus, posztkoordinált), az internetkatalógusok pedig a szintetikus (generalizáló, hierarchikus, prekoordinált) tartalmi feldolgozást és keresést teszik lehetővé.

1876-ban, az amerikai könyvtártörténet „csodálatos évében” ugyancsak szinte egy időben született meg az a két rendszer, amely lényegében alapja lett a modern osztályozásnak és információkeresésnek. Dewey Tizedes Osztályozása és folytatása, az ETO a Gopherhez hasonlóan hierarchikus szervezettségű volt, és a ráépülő

szakkatalógusok ugyancsak fölfoghatóak egyetemes léptékű, korlátozott számú csúcsfogalommal rendelkező óriási tartalomjegyzékeknek, mint a mai internetkatalógusok. (Az egyik – eredetileg Gopher-menüként született – szolgáltatás, az 1989-ben született CyberDewey nevében is utalt erre a rokonságra.) Az internetkatalógusok is az ETO szellemiségén alapulnak: meghatározott, könnyen áttekinthető számú főosztály, és a belőlük kiinduló alosztályok alá-fölé rendeltségi szerkezete jellemzi őket.

Ezzel szemben Cutter természetes nyelven alapuló tárgyszórendszerének elvileg végtelen sok hierarchiacsúcsa lehet, akárcsak a web indexelő rendszereinek. Az összetett tárgyszavakon belül ugyan érvényesült kezdetben valamiféle hierarchikus szervezettség, de a tárgyszavakat mind szabadabban kezdték egymással kombinálni, és idővel az összetett tárgyszavak használatától is eltekintettek. Fokozatosan kialakultak – Taube „uniterm” rendszerének hatására – a deskriptoros információkereső nyelvek és szótáraik, a tezaurusok. Szerkezeti felépítésük nagyon emlékeztet a hipertexten belüli kapcsolódásokra, mert a tezaurusok deskriptorai tetszés szerinti deskriptorral összekapcsolhatók és fordítva (azaz a kapcsolódások szerkezete itt is M:N, akárcsak a hipertext esetén). Az indexelőszolgáltatásokban az egyedi szavakat tartalmazó indexek alapján végezhető a természetes nyelven alapuló keresés, és egyre gyakoribb, hogy ezt kötött keresőszótárak, tezaurusok alkalmazásával támogatják.¹²

4.1.4 A szerver- és kliensoldali keresés

A születés lázában és nem utolsósorban a nagyobb hírverés kedvéért olykor hangzatos, olykor meg többjelentésű megnevezéssel találkozunk a felhasználó.

A navigálás szempontjából az internetet – William Gibson 1984-ben írt Neuromancer című fantasztikus elbeszélése nyomán [6] – afféle virtuális „kibertérnek” (cyber space) nevezik. E „térben” a weben folytatott kereséshez külön szoftvereket használnak a szerver-, és külön szoftvereket a kliensgépeken, melyeket szerver-, illetve kliensoldali „navigációs eszközöknek”, egyszerűbben szerver-, illetve kliensprogramoknak is neveznek. (A még egyszerűbb tolvajnyelvi „kliens” és „szerver” elnevezést a bennfentesek bizonyára a kezdők elriasztására használják, nehogy egyértelmű legyen számukra, mikor van szó gépről, és mikor programról.)

Meg kell különböztetni a kétfajta keresőszolgáltatáshoz szükséges szerveroldali és kliensoldali keresőeszközöket. A szolgáltatói vagy szerveroldalon található az indexelőszolgáltatások és

az internetkatalógusok (böngészőszolgáltatások) programjai és adatbázisai, a felhasználói vagy kliensoldalon pedig visszakereső („nézegető”, viewer, browser) rendszerek.

Az indexelőszolgáltatások szervergépein a következő szoftvereket használják:

- a „leszedőnek” (krawler, crawler, spider, wanderer, gatherer, scooter) vagy „robotnak” nevezett szoftver, amely afféle webvándorként járkal a kibertérben, és a HTML-dokumentumokon beágyazott hipertext csatolókat kihasználva mozog egyik dokumentumról a másikra, hogy felhasználói beavatkozás nélkül egyetlen mutatóba gyűjtse össze a HTML-dokumentumok kulcsszavait;
- ezt egészíti ki az indexelő- (indexkészítő) szoftver (pl. WAIS, Archie, INQUERY, Apache, Glimpse, Harvest), amely a felkutatott, indexelt HTML-dokumentumok másodlagos adatait (leírásait) adatbázisokba rendezi. Az indexkifejezéseket automatikusan generálják, különös figyelemmel a HTML-dokumentumok címfejlében szereplő másodlagos (meta-)adatokra.

Az adatbázis tartalma az URL, kulcsszavak, webcím, rövid tartalmi kivonat, teljes szöveg első sora stb. Ezek alkotják a másodlagos információteleket vagy rekordokat, az indexelt HTML-dokumentumok pedig az elsődleges dokumentumok. (A másodlagos információkat hálózati és digitális könyvtári környezetben metaadatoknak nevezik, az elsődleges információkat pedig a digitális objektumoknak.)¹³

Az adatbázisban tárolt információk frissítése kumulatív vagy reprodukáló szokott lenni. Az előbbi esetben az új rekordok hozzáadódnak a meglévőkhöz, az utóbbi esetben pedig időközönként az új rekordokkal a teljes adatbázist újjászervezik (rekordon itt a keresőszolgáltatások által összegyűjtött másodlagos információteleket értjük, elsősorban HTML-dokumentumok leírásait);

- a „leszedő” és az indexelőprogramot integráló egységet, amely egyben elvégzi a kliensoldalról közvetített szerveroldali keresést is, összefoglalóan „keresőgépek” („keresőmotor”, „keresőmű”, „search engine”), szerényebb megnevezéssel keresőrendszernek (search system) nevezik.

Előfordul, hogy „keresőgépen” nem szerveroldali szoftvert, hanem azt a számítógépet értik, melyet a keresőrendszer futtatására állítottak üzembe. Még gyakoribb, hogy magát a szerveroldali teljes keresőszolgáltatást nevezik „keresőgépek”, „keresőmotor”, „robotnak” (search engines, bots, robots, Suchmaschinen, Roboter), noha e megnevezések csak a szerveroldali leszedő, indexelő és kereső programrendszerek együttesét, robot esetében pedig a leszedőt jelölik, és nem a teljes szolgáltató rendszert, melybe beletartozik még a felhasználói felület és a szolgáltatott tartalom is. (Azt mondják, hogy az AltaVista „keresőgép”, holott az AltaVista a teljes keresőszolgáltatás neve, melyen belül – többek között – leszedők, indexelő- és keresőprogramok működnek.)

Az internetkatalógusok (böngészőszolgáltatások) szervergépein a következő szoftvereket használják:

- a ma még szinte kizárólag intellektuálisan osztályozott HTML- oldalak másodlagos adatait (leírásait) kezelő adatbázis-kezelő rendszer, amelybe az alkalmazott rendező rendszert is integrálták (ez tehát nem indexelést végző „keresőgép”, noha indexelő-programok kiegészítő alkalmazása is mind gyakori);
- szükség esetén a felhasználóbarát megjelenítést biztosító előtétprogram.

A felhasználói kliensgépeken – a kliensoldalon – fut az ablakos, felhasználóbarát vizuális felülettel ellátott „nézegetőnek”, „böngészőnek” (viewer, browser) nevezett szoftver (mint amilyen például a Mosaic, a Netscape Navigator, az Internet Explorer stb.). Ezek a programok valójában nem keresnek, hanem a felhasználó által kijelölt keresési parancsokat közvetítik a szerveroldali automatikus keresést elvégző „keresőgépek”, ezért is neveztük őket feljebb visszakereső rendszereknek, mivel egyszer már kikeresett rekordokra irányulnak. Hívják őket közvetítőeknek is.

A szerveroldali szoftverekkel a felhasználó mindig csak a kliensoldali „nézegetőn” keresztül kerül kapcsolatba. A „keresőgépek”, adatbázis-kezelők stb. a háttérben automatikusan működnek.

A szerveroldali szoftverek tehát olyan kliensoldali szoftvereket igényelnek, amelyekkel az előbbiek szolgáltatásai realizálhatók a felhasználóknak. Az utóbbiak fogadják a keresőkérdéseket, megteremtik az összeköttetést a szerverprogramokkal, és ezeknek a választát megfelelően „kiszerve” közvetítik a felhasználónak. Ezt az egymást feltételező szoftverszerkezetet nevezik kliens-szerver rendszernek.

A nézegetők mindinkább az internet felhasználói rendszerei lesznek. Nekik köszönhető, hogy a nagy jelentőségű, ám nehezen hasznosítható számítógépes kapcsolatokról informatív és könnyen kezelhető kommunikációs eszköz lett. A végfelhasználó nem is veszi észre, hogy a nézegető használatkor indexelőfolyamat eredményében részesül, mert eltakarja előle a felhasználóbarát, „természetelvű” felhasználói felület.

4.2 Indexelőszolgáltatások („keresőgépek”)

4.2.1 Meghatározás

Az indexelőszolgáltatások „keresőgépeket” alkalmazó szolgáltatások (robot generated indices), melyek adatbázisa a „keresőgépek” által indexelt HTML-dokumentumok rekordjait (másodlagos adatokból álló leírásait) tartalmazza. Bennük természetes nyelvű szavakkal végezhető a lekérdezés.

Az ismertebb globális rendszerek közé tartozik például az AltaVista, Excite Search, HotBot,

Infoseek, Lycos A2Z, Northern Light. A magyarországi webhelyeket 1996 óta a Heuréka (Hungary. Network) dolgozza föl, 1998 után pedig megjelent az AltaVista (Matáv) is.

A szolgáltatások leszedői éjjel-nappali üzemben, csatolóról csatolóra haladva indexelik a HTML-dokumentumokat. Jelentős részük a teljes szöveget indexeli, de közülük sokan a teljes szövegből csak meghatározott számú sort (pl. az első húsz sort) és a metaadatokat veszik figyelembe (pl. a Lycos A2Z). Léteznek szolgáltatások, amelyek eleve csak a HTML-dokumentumok metaadatait vagy kis részüket dolgozzák föl (pl. a WWW Worm).

Az egész szolgáltatást hibásan „keresőgépnak” nevezik, holott a „keresőgép” a szolgáltatásnak csak egyik része.

4.2.2 Indexelés, „begyűjtés”

Az indexelőszolgáltatások fontos jellemzője a gyűjtőkör és a kiválasztási-indexelési módszer. Az elsőre szerencsés esetben már a szolgáltatás nevéből következtetni lehet, és mindig található a belépőlapra olyan csatoló (pl. Magunkról, About Lycos), amelyet működtetve a szolgáltatás céljáról tájékozódhatunk. A másodíkról csak közvetett információk állnak rendelkezésre, a belépőlapról kiindulva e tekintetben semmiféle érdemleges adathoz nem lehet jutni. Számos vizsgálat a szolgáltatások közvetlen megkeresésével készül el.

„Az adott »begyűjtési stratégia« (gathering, harvesting) és forrásfelkutatás (resource discovery) dönti el, hogy milyen szervereket talál meg a keresőgép, és azon belül milyen dokumentumok indexelését részesíti előnyben. Az indexelt egységek száma szolgáltatásonként különböző, néhány tízezer (Harvest Home Page Broker) a tizen- és huszonmillióig terjed (AltaVista, Lycos, Northern Light). De hogy mit tekintenek egységnek, az ugyancsak szolgáltatásonként változó. Van, amelyik – mint a Lycos – minden elért URL-t számol, noha a dokumentumoknak csak töredékét indexeli, az Open Text annyiszor számolja az URL-t, ahányszor az a legkülönbözőbb dokumentumokban előfordul, az Inktomi viszont csak a teljes szövegükben indexelt dokumentumokat számolja.

Az »először átfogóan« (breadth-first) indexelő stratégiát alkalmazó rendszerek gyűjtőköre nyilván nagy lesz, az »először mélyen« (depth-first) indexelő stratégia eredménye pedig a részletesen indexelt, de kevés dokumentum, egyben kevés begyűjtött szerver lesz” [13].

A szervergépen kezelt adatbázisba betárolt adatok az indexek alapján kérdezhetők le a kliensoldali nézegetőkkel. A találatokat elemzik, és többnyire relevanciavizsgálatnak is alávetik. Az

indexelőszolgáltatások szempontjából a HTML-dokumentumok internetforrások, és a HTML-dokumentumok meghatározott összessége a keresőszolgáltatások „gyűjtőköre”.

Vannak olyan keresőszolgáltatások is, amelyek katalógusokból (is) készítenek indexeket a lekérdezéshez (pl. ALIWEB, Yahoo! Search¹⁴, InterCat), és számos kereskedelmi szolgáltatáshoz ingyen be lehet jelentkezni.

„Az indexelőszolgáltatásoknak be is lehet küldeni HTML-dokumentumokat, amit szívesen vesznek, mert bővíti a választékot. (Az internetkatalógusok kisebbik része kizárólag ezen az alapon működik.) A manuálisan gyűjtött, intellektuálisan feldolgozott indexek előnye a tartalmi ellenőrzöttségben rejlik. Olyan tételek indexei ezek, amelyeket vagy a szolgáltatás szakembere, vagy a szerző maga dolgozott föl. Hiába állnak rendelkezésre jól szerkesztett bejelentkezési űrlapok, pl. az ALIWEB tapasztalatai alapján ezeket többnyire felületesen töltik ki. A nem szöveges dokumentumok esetén nyilván mindig szabványosított bevételi űrlapokat kell használni” [13].

4.2.3 Avulás és frissítés

A HTML-rekordok hamar avulnak, mert a HTML-dokumentumok megszűnhetnek, átalakulhatnak. A feldolgozott állomány frissítése az indexelő rendszerek többségében elvileg könnyebben megoldható, mint az internetkatalógusokban, ahol intellektuálisan osztályozzák a HTML-dokumentumokat, és az automatikus frissítés hiányában kialakulnak a zsákutcás, halott tételek (dead links) [18].

„A keresőgépeken alapuló szolgáltatások legnagyobb előnye, hogy a körülményekhez képest rendkívül gazdagok. Mivel az esetek többségében az eredeti források lényeges részeit, sokszor a teljes szöveget indexelik, nagy a valószínűsége annak, hogy rendkívül speciális információk is megtalálhatók. Éppen ez a tény indokolja, hogy előbb-utóbb érdemes lesz jobb eszközöket is rendelkezésre bocsátani az információkereső stratégiához.

Az aktualizálás gyakorisága a hetenkénti (pl. Lycos, Webcrawler), a félévenkénti, sőt évente egyszeri gyakoriság (WWW Worm) között mozog. A legtöbb szolgáltatás nem közöl erről semmit. Ugyanannak a HTML-dokumentumnak különféle változataiból közelítőleg megállapíthatók az erre vonatkozó adatok. Mennél nagyobb a szolgáltatás, annál kisebb frissítési gyakoriságra lehet számítani” [13].

Az elemzések szerint a nagyobb szolgáltatások többségében a halott tételek száma megközelíti a 20–30%-ot. A kisebb szolgáltatásokban a helyzet ennél lényegesen jobb (lásd az 1. táblázatot).

1. táblázat

Lekérdezés eredményeként kapott nem élő tételek száma 1999. 03. 05-én [18]

Keresőszolgálat	Nem élő tételek %-a
Lycos	29
AltaVista	18
Northern Light	16
MSN Web Search	14
Yahoo!, Inktomi	13
Snap!	11
Infoseek	8
HotBot	4
Google!	0
Excite	0

Elvileg az internetkatalógusokban is automatikusan elvégezhető volna a frissítés azáltal, hogy megfelelő program törli a már nem előhívható HTML-dokumentumok kapcsolatait az indexadatokhoz, de erről nincs információ.

A mennyiségi teljesítmények lenyűgözőek: az AltaVista keresőrendszerében pl. naponta kb. 10 millió HTML-dokumentumot néz át a leszedő, ez közel tizede a több mint 128 millió indexelt tételnek, amelyet a rendszer adatbázisa tartalmaz.¹⁵ A 2. táblázatban néhány keresőszolgálat adatbázisának mérete látható.

2. táblázat

Keresőszolgáltatások adatbázisainak mérete 1999. 03. 05-én [18]

Keresőszolgálat	Rekordok száma
Northern Light	128 540 264
AltaVista	106 169 808
HotBot/Anzwers	99 409 035
Snap!	98 638 820
Google!	71 065 137
Infoseek	59 700 192
MSN Web Search	39 589 032
Excite	32 896 723
Lycos	22 781 237

4.2.4 Keresési módszerek és stratégia

A szolgáltatások általában arra törekszenek, hogy a teljesség (recall) legyen nagy, ezért pontosságról eleve nincs szó. Az alkalmazható módszerek, stratégia meglehetősen változatosak. Az alapértelmezésen túlmenő lehetőségek (részletes keresés, advanced search) a szolgáltatások jelentős részénél nincsenek előtérben, a laikus sokszor nem is veszi észre őket.

- A kereséshez egyedi szavakat adhatunk meg.
- Hozzáértők választhatnak más Boole-operátorokat és helyzeti (távolsági/közelségi) operátorokat.

- Alkalmazható a „szólánccal” végzett keresés (string-search), hol idézőjelek közé téve a láncot, hol legördülő mezőn minősítve.
- Többnyire megadható, hogy csonkoltan vagy pontosan értelmezendő-e a keresőszó. Egyes rendszerekben (pl. AltaVista) megkülönböztethető a kis- és nagybetű.
- A szolgáltatások kis részénél (AltaVista, Excite Search, Lycos A2Z) megadható a nyelvi, sőt – mint a Lycos esetében – néhány dokumentumtípus szerinti szűkítés is, azaz kérhető csak meghatározott nyelvű vagy dokumentumtípusba tartozó találatok megjelenítése.
- Néhány szolgáltatás, mint pl. az AltaVista „idézetes” keresést (citation indexing) is lehetővé tesz, azaz megadja azokat az összetett szavakat, amelyekben az egyedi keresőszó előfordul, és ezeket fölhasználva szűkíthető a keresés (pl. a „műanyag” kifejezéssel keresve felajánlja a „hőre lágyuló műanyagok”, „ipari műanyag burkolatok” stb. kifejezéseket is a kereséshez).
- Van olyan szolgáltatás, amelyben kiköthető, hogy a keresés csak a HTML-címben, az összefoglalásban vagy a teljes szövegben szereplő szavakra korlátozódjék, vagy elvéve kiköthető, hogy a dokumentumból mely oldalak jelenjenek meg.
- Van olyan szolgáltatás (Highway 61), amelyben megadható, hogy a kereső milyen színvonalú lekérdezést igényel.
- Olykor az elvégzett keresés eredményhalmazán végezhető másodlagos keresés (relevanz feedback, find similar pages). Ez annyit jelent, hogy a találat-hoz lekérhető az adott találat-hoz „hasonló” tartalmú (similar, related topic, Wortverwandschaft) tételek.
- Az Infoseek „specifikus keresést” is biztosít: az átfogó jelentésű kifejezéshez a | (vonat, pipe) jellel megadható a specifikus (pl. „tánc | tangó” esetén a „tánc” alapján kiválasztott halmazból a „tangóval” jellemzett rekordokat kapjuk meg).
- A fejlettebb rendszerekben (pl. MetaGer) az is kérhető, hogy ellenőrizték, élnek-e még egyáltalán a talált tételek? Ilyenkor a végeredményre valamivel tovább kell várni.

Az egyik nagy probléma a keresőszolgáltatások túlnyomó részében, hogy a keresés nem korlátozható mezőkre (formátumszegmensekre), és a találatok csak néhány szolgáltatásban rendezhetők különféle szempontok (dátum, hely stb.) szerint. Létezik néhány kivétel: a Lycosban például kiköthető, hogy a keresés a teljes szövegben, a címben vagy az URL-azonosítóban történjék; az Infoseek lehetővé teszi a találatok rendezését dátum szerint is. A viszonylag már elterjedt nyelvi szűkítés mellett olykor a regionális finomítás is lehetséges (Yahoo! Get Local).

A keresési végfelület (ablak) a legtöbb szolgáltatásban a végletekig egyszerű, általában semmiféle keresési segítséget nem tartalmaz. Ehhez a megfelelő, alig észrevehető csatolót kell megkeresni (részletes keresés, advanced search).

Csak kevés szolgáltatás adja meg az eszközök választékát az első oldalon. Noha általában a „legostobább felhasználóra” számítanak, ehhez képest a keresési segítség, különösen pedig a keresési példák hallatlanul szegényesek, az online szolgáltatásokban természetes keresési stratégia és keresőkép (profil) fogalmait teljesen ismeretlenek – legalábbis egyelőre.

A keresés finomítása terén az internet indexelőszolgáltatásai általában még alulmaradnak az online szolgáltatásokkal szemben. Az utóbbiakban az alkalmazott hagyományos adatbázis-kezelő rendszerek jóvoltából az információkeresési stratégia teljes tárháza rendelkezésre áll. Ez a helyzet azonban rohamosan változik. A felhasználói komfort dolgában az indexelőszolgáltatások már ma nem egy vonatkozásban előbbre vannak.

Az üzleti szempontok következtében egész sor tájékoztatási komforttal látják el a felhasználót. Ilyen például a leggyakrabban használt keresőszavak százalékban megadott gyakorisága. Első helyen ugyan az erotikus információk keresettségére utaló kifejezések állnak, sokkal jelentősebb azonban, hogy ezután a közhasznú dolgokra (állás- és társkeresés, közintézmények, adattárak, telefonkönyvek, menetrendek) vonatkozó keresőszavak következnek. Ezt követik a rendkívül kis gyakoriságú speciális szakkifejezések (az „epitaxiától” az „aloe veráig”, az elméleti matematika kifejezéseitől a teológiai fogalmakig). Mindebben az a fontos, hogy gyakorlatilag a szaknyelv minden elképzelhető szavát használva tesznek föl keresőkérdéseket. A komoly keresők részéről tehát rendkívül nagy és differenciált igények jelentek meg. A 3. táblázatban ilyen gyakorisági jegyzék nagyon leegyszerűsített kivonata látható.

3. táblázat

Kivonat az AltaVizslában megadott felhasználói keresőszavak gyakorisági jegyzékéből 1998 májusa és augusztusa között

Keresőszó	Gyakoriság (%)
szex	8,91
pornó	2,38
erotika	2,19
magyar	0,98
társkereső	0,55
társkeres	0,54
telefonkönyv	0,49
önkormányzat	0,11
tenzor algebra	0,0098
szikraforgácsolás	0,0022
kontrakció	0,0019
tűlhűlés	0,0017
Pragmatica Sanctio	0,0008

4.2.5 Találatmegjelenítés

A megjelenített másodlagos információtétele (HTML-dokumentum találati leírása) többnyire egyszerű, és szolgáltatásonként különbözik. Nincs szabványosított megjelenítési forma (egységes megjelenítés legfeljebb az itt nem tárgyalt online szolgáltatásokban fordul elő, de ott is nagyon ritkán felel meg bibliográfiai szabványoknak). Sokszor megadható, hogy egyszerre hány találat jelenjék meg, nagyon kevés rendszerben (pl. Infoseek) lehetséges nemcsak relevancia, hanem dátum szerint is rendezni.

A 2. ábrán a Heuréka és az AltaVizsla keresőszolgáltatások egyszerű információtételei láthatók.

Ritka kivétel az olyan szolgáltatás, mint a Lycos, amelyben a találatok leírása, azaz a másodlagos információtétele részletes és gondosan strukturált (3. ábra).

Relevancia mértéke Címfej Szövegkezdet URL-azonosító Méret, dátum

[77%] **Szakadát István**
 Szakadát István: Xanadu Xanadu és Ted Nelson. A Xanadu egy már több mint 30 éve tartó informatikai projekt Nelson szakmája egyik sokat vitato...
<http://www.univworld.hu/netskills/tudas/HTML/Xanadu.htm> - hossz: 39468 - 99 Apr 14

1. Szakadát István
 Szakadát István: Xanadu. Xanadu és Ted Nelson. A Xanadu egy már több mint 30 éve tartó informatikai projekt, Nelson szakmája egyik sokat vitatott...
 ★★★★★ | <http://www.univworld.hu/netskills/tudas/HTML/Xanadu.htm> | méret: 39468 byte | dátum: 18-Dec-1998

2. ábra Ugyanannak a találatnak információtételei (HTML-rekordjai) a Heuréka, illetve az AltaVizsla keresőszolgáltatásokban

Book Search Results

Here are your search results for Keyword is **Internet**

We found 4,457 matching titles.

1 - 25 are displayed below in **bestselling order**.

We also found 1,212 Out-of-Print and Used Titles

Re-sort my search in: **Bestselling**, A to Z, Date Published order.

1. **HTML 4 for the World Wide Web: Visual QuickStart Guide**

In-Stock: Ships within 24 hours.

Elizabeth Castro, Nancy Davis (Editor) / Paperback / Date Published: January 1998

Retail Price: \$17.95 Our Price: \$12.56, You Save \$5.39 (30%)

► [Buy this book or read more about it](#)

2. **The Internet for Dummies**

In-Stock: Ships within 24 hours.

John R. Levine, Margaret Levine Young, Carol Baroudi / Paperback / Date Published: January 1999

Retail Price: \$19.99 Our Price: \$15.99, You Save \$4.00 (20%)

► [Buy this book or read more about it](#)

3. ábra A Lycos részletes találati leírásai

4.2.6 A találatok relevanciája

A relevancia mértékét egyrészt annak alapján állapítják meg, hogy a keresőszó a HTML-formátum címfejében (<Title>) szerepel-e, vagy csak a tartalmi kivonatban, illetve szövegben, és az utóbbin belül milyen gyakorisággal. Másrészt automatikus indexelési módszereket használnak. Az ismertebb indexelőprogramok közül pl. a WAIS a vektortérmodellt alkalmazza: az indexelt kifejezések alapján dokumentumvektort számít ki, és ezt hasonlítja össze a keresőkérdés vektorával, kiszámítva a kettő közötti távolságot egy n-dimenziós vektortérben. Mennél kisebb a távolság, annál nagyobb a relevancia. Az INQUERY az interferencia-hálók modelljét használja: e hálók a keresési folyamat határozatlanságát képzik le, melyből valószínűségi módszerek segítségével számítják ki a relevanciát [20]. Ahhoz képest, hogy a dokumentációs célú automatikus indexelés és osztályozás terén néhány évtizeddel korábban milyen eredmények születtek, a keresőszolgáltatások relevanciavizsgálatai – egyelőre még – meglehetősen szegényesek.

A talált tételek rendkívül vegyes minőségűek. Gyakori, hogy már nincs is mögöttük élő tartalom-szolgáltatás, sokszor ugyanannak a HTML-

dokumentumnak különböző időpontokból származó változata jelenik meg, és a találatok túlnyomó többsége valójában teljesen irreleváns, mivel az indexelt szó nem a HTML-dokumentum tartalmát reprezentálja. Általában elmondható, hogy ha a relevancia mértéke az 50% alá csökken, a találat már teljesen irreleváns.

Olykor maguk a kereskedelmi tartalom-szolgáltatók is tovább rontják a találatok minőségét: HTML-dokumentumaik címfejébe olyan kifejezéseket is elhelyeznek, melyek valójában nem igazán jellemzik a lapjukat, de amelyekről tudják, hogy a gyakran keresettek közé tartoznak, hogy a leszedő – rájuk találva – különlegesen értékes találatként értékelje, és a találati jegyzéken a legelső helyeken jelenítse meg őket.

Az indexelőszolgáltatások leszedőinek túlnyomó többsége kötött szótár nélkül válogatja ki a szövegszavakat (a hagyományos információkeresés nyelvén ez „szabad szövegen belüli keresés”, „szabad kulcsszavas keresés”). Ritkán előfordulnak közöttük olyanok, amelyekben kötött szótárat, néhányukban (pl. AltaVista, EEL, Kolibri) tezausztruszt is használnak. Ez utóbbi szolgáltatások információtételeinek relevanciája általában lényegesen nagyobb.

4.3 Gyűjtő- és többszörösen indexelő szolgáltatások

Az indexelőszolgáltatások választéka ma már rendkívül nagy. A közel tucatnyi nemzetközileg ismert rendszeren kívül nagyon sok a speciális gyűjtőköri (dedikált) rendszer, amely csak meghatározott típusú HTML-dokumentumokat dolgoz föl (pl. könyvkiadók legújabb kiadványait, műszaki folyóiratokat, folyóiratcikkeket, Usenet-cikkeket, cégeket), továbbá az olyan rendszer, amely csak meghatározott államon belüli webhelyek HTML-dokumentumait indexeli (a Heuréka és az Alta-Vizsla pl. csak a magyarországiakét). A szolgáltatások teljesítménye kisebb-nagyobb mértékben különbözik egymástól, ami a felhasználót arra kényszeríti, hogy a lehető legnagyobb teljesség érdekében több indexelő keresőszolgáltatást is igénybe vegyen, ami meghosszabbítja a keresést.

Ráadásul nehéz összehasonlítani az eredményt, mert a találatok külön-külön jegyzékekben jelennek meg.

Ezen hivatottak segíteni a *többszörösen, szimultán vagy meta-keresőszolgáltatások* (*multiple/parallel/meta search engines, Meta-Suchmaschinen*). Velük egyszerre több indexelő keresőszolgáltatásban lehet keresni anélkül, hogy a felhasználónak az egyes szolgáltatásokkal külön foglalkoznia kellene. Ez körülbelül olyan, mintha valaki könyvet keres, mégpedig az összes magyarországi könyvtárban, és megkapja találatként, hogy az adott könyv milyen adatok kíséretében található meg az egyes könyvtárakban.

A többszörösen indexelő szolgáltatás leszedője a kijelölt indexelőszolgáltatásokat a keresőkérdekek alapján párhuzamosan fésüli át, és a találatokat közös listában jeleníti meg, ami a nagyobb választék mellett a jobb összehasonlítást is előse-

Geben Sie einfach ein oder mehrere Suchwörter ein:

Thesaurus

Alle Wörter sollen im Dokument vorkommen ☐

☐ mit internationaler Suche, Metacrawler

☐ NEU: Ausgabe alphabetisch nach Servern zusammenfassen

☒ mit MetaGer QuickTips ... und Sprücheklopfer: ☒

☒ Treffer bei Anklicken in neuem Fenster öffnen

20 Sekunden maximale anfängliche Suchzeit

☒ keine Linküberprüfung

☐ Teste Existenz und sortiere aktuellste zuerst

☐ Teste Existenz und sortiere nach Relevanz

Die von uns ausgewählten besten Suchdienste:
(Falls Sie Voreinstellungen unter diesen ausgewählten Diensten ändern wollen, klicken Sie bitte den entsprechenden Schalter an.)

☒ Netfind	☒ Altavista	☒ Infoseek	☒ yahoo.de	☒ witch
☐ Speedfind	☒ Netguide	☒ Lycos	☒ T-Online	☒ Crawler.de
☐ MSN	☐ de.*-News	☐ Dino	☐ Fixx	☐ Kolibri



4. ábra Meta-keresőszolgáltatás belépőlapja

A „Quick Tip” (más rendszerekben a „Direct Hit”) azokat a külön megjelenő találatokat adja meg, melyek a keresőszót az URL-névben tartalmazzák. Kérhető a találatok egzisztenciájának ellenőrzése, és a dátum vagy a relevancia szerinti rendezés. A táblázatban felsorolt, egyszerre lekérdezhető keresőszolgáltatások választéka, beállítása megváltoztatható, és kérhető, hogy a keresésben az angol „MetaCrawler” is részt vegyen.

gíti. Az első ilyen rendszerek 1995-ben készültek. Hatékony működésüket az elosztott rendszerű szervezéssel fokozták: a részműveletekre felbontott feladatokat egyszerre több számítógép leszedő- és indexelőprogramjaira delegálják. A legismertebb ilyen többszörös feladatmegosztásra képes rendszer – a Harvest –, amely a legfejlettebbek közé tartozik, jelenleg már ingyen hozzáférhető.¹⁶ Számos ismert indexelőszolgáltatásban használják.

Elég megadni a keresőkérdést, a többszörösen indexelő szolgáltatás a profiljába fölvetett szolgáltatásokat végignézzve kilistázza a találatokat. A jobb minőségű rendszerekben a talált információ-tételeknél feltüntetik, hogy melyik indexelőszolgáltatásból származnak. Így gyakran ugyanaz a tétel többször is megjelenik, és összehasonlítható, hogy melyik rendszer szolgáltatja a legfrissebb találatokat. A 4. ábrán a „MetaGer” keresőszolgáltatás belépőlapja látható (az angol „MetaCrawler” belépőlapja nem annyira strukturált és informatív, ezért választottuk a német változatot).

Mára kialakultak ezeknek a szolgáltatásoknak a kritériumai:

- *párhuzamos keresés*, azaz egyszerre több szolgáltatás lekérdezése egyetlen keresési műveletben;
- *eredmény-összefésülés*, azaz a találatok megjelenítése egyetlen formátumban;
- *többszöröződések kezelése*, azaz ugyanazt a HTML-dokumentumot a rendszernek fel kell ismernie, és jelölnie kell az egyes forrásokat, amelyből származik;
- *ÉS- meg VAGY-művelet* mint minimális logikai keresési eszköz;
- *információvesztés nélküli működés* (ha pl. az egyik forrás tartalmi kivonatokat tartalmaz, azt át kell tudni venni);
- *forrásrendszer-elfedés* [hiding] (a lekérdezett indexelőszolgáltatások tulajdonságai nem játszhatnak semmiféle szerepet a metarendszer szintjén, a felhasználónak semmit sem kell tudnia ezekről a specifikumokról);
- *teljesség* (a keresésnek addig kell tartania, ameddig a lekérdezett szolgáltatásokból találatok nyerhetők).

Roget's Thesaurus

DirectHit: Roget's **Thesaurus** of English Words and Phrases, Classified and Arranged so as to Facilitate the Expression of Ideas and Assist in Literary Composition.

Infoseek: Roget's **Thesaurus** of English Words and Phrases, classified and arranged so as to facilitate the expression of ideas and assist in literary composition.

LookSmart: ...

1000, <http://www.thesaurus.com/> (*Direct Hit*, *Infoseek*, *LookSmart*, *Thunderstone*, *Yahoo!*, *GoTo.com* paid result)

ARTFL Project: ROGET Form

DirectHit: 8/18/99: Please note that the server is currently down due to hacker activities. We are making every effort to restore the system as quickly as possible. New!! The ARTFL Project

Infoseek: Searchable text of "Roget's **Thesaurus**," version 1.02.

WebCrawler: 8/18/99: Please note that the server is currently down due to hacker activities. We are making every effort to restore the system as quickly as possible.

625, http://humanities.uchicago.edu/forms_unrest/ROGET.html (*Direct Hit*, *Infoseek*, *WebCrawler*)

NASA Thesaurus

DirectHit: contains the authorized subject terms by which the documents in the NASA STI Databases are indexed and retrieved. The NASA **Thesaurus** comprises two volumes: Volume 1 – Hierarchical

Infoseek: Look up words about space, flight, technology and more. Updated daily.

WebCrawler: ** This Web Site has been superseded ** You will be automatically redirected to the replacement site (<http://www.sti.nasa.gov/thesfrm1.htm>). Please update your site links, bookmarks, or indexes.

547, <http://www.sti.nasa.gov/nasa-thesaurus.html> (*Direct Hit*, *Infoseek*, *WebCrawler*)

Pages at this site with the same name:

<http://www.sti.nasa.gov/nasa-thesaurus.html>

5. ábra A találatok megjelenítése a MetaCrawler meta-keresőszolgáltatásban

A keresőkérdés „Thesaurus” volt. A legelső sor elején a relevancia mértéke látható (az első találatot a rendszer 100%-osnak értékelte), és a sor végén szerepelnek mindazok az egyedi keresőszolgáltatások, melyekben a találat szerepel.

Robot generated WWW Indices

These indices of WWW-based resources are generated by robots, and therefore very complete, but are more likely too much information.

Search term:

Lycos

WebCrawler

ElNet Galaxy

Harvest WWW Home Pages

InfoSeek Demo

Open Text Corporation

JumpStation

RBSE

Nikos

The Lycos robot is probably the largest.
 The WebCrawler is smaller, but more up-to-date.
 The comprehensive Harvest Resource Discovery s
 The ElNet Galaxy also has a subject tree.
InfoSeek is commercial, but have a limited free de
JumpStation is an index in the UK, which is not ac
 If they don't help, the RBSE URL Search, the Nikos,

You might also try Open Text, who cannot currently be queried via CUSI.

Other Internet Indices

These are not WWW-based, but may well be of interest.

6. ábra Gyűjtőszolgáltatás oldalának részlete, melyen a „keresőgépeket” használó keresőszolgáltatások szerepelnek

Az éppen legördített jegyzékből választható ki, hogy a keresőkifejezést melyikben kívánjuk lekérdezni. Az aláhúzott csatolókat működtetve át lehet térni az egyes szolgáltatások belépőlapjára is.

E követelmények együttes kielégítése minden jel szerint komoly nehézségekkel jár, mivel jelenleg világszerte alig néhány metarendszer képes teljesíteni (MetaCrawler, MetaGer és Highway 61). Az 5. ábrán a MetaCrawler találatai láthatók.

Egyszerűbb, átmeneti típus a gyűjtőszolgáltatás (configurable unified serach interface [CUSI], all-in-one formular, sample service, Sammeldienst), amely felkínál több, olykor nagyon sok keresőszolgáltatást (mintegy gyűjteményként), de mindig csak egyet lehet kiválasztani a lekérdezésre. (Ilyenek például az 1. ábrán látható BSZ, a CUSI különféle regionális változatai, Magyarországon pedig a HUDIR-t és a Heurékát fenntartó Hungary. Network gyűjtőszolgáltatása: „Internetkeresés”).

A felkínált szolgáltatásokat sokszor rövid magyarázattal is ellátják és csoportosítják; mindig csak az azonos típusba tartozó csoport valamelyik tagja kérdezhető le (6. ábra).

A többszörös indexelőszolgáltatásokhoz hasonló, de más szervezettségű például a karlsruhei könyvtár internetkatalógusa (Karlsruher Virtueller Katalog), melyen keresztül az összes nagyobb német – valamint a Kongresszusi, az angol, francia stb. nemzeti – könyvtár állománya egyetlen keresőkérdelemmel elérhető.

Magyarországon a Közös Elektronikus Katalógusban (KözElKat), a Voyager adatbázis-kezelő rendszert (VOCAL) és a TINLIB adatbázis-kezelő rendszert (AROMO) használó könyvtárak közös katalógusában lehet egyetlen felületen a részt vevő könyvtárak állományában elvégezni a keresést.¹⁷ Akárcsak a karlsruhei esetben, ezek sem (szűkebb értelemben vett) keresőszolgáltatások, hanem távolsági online adatbázis-hozzáférést biztosító szolgáltatások összevont, közös webfelületen (ún. „brókeren”).

Irodalom

- [1] BUSH, V.: As we may think. = Atlantic Monthly, 176. sz. 1945. júl. p. 101–108. Magyarul: Út az új gondolkodás felé. = Hypertext + multimédia. Oktatási segédanyag. Budapest, Artpool, 1996. p. 3–14.
- [2] BUSH, V.: Endless horizons. Washington, Public Affairs Press, 1946.
 BUSH, V.: Memex revised. = Sciences is Not Enough. New York, William Morrow, 1967. p. 75–101. <<http://www.csi.uottawa.ca/~dduchier/misc/vbush/awmt.html>>

- [3] DESIRE (Development of a European Service of Information on Research and Information). EU Project, 1996. <<http://www.ub2.lu.se/desire>>
- [4] DRÓTOS L.: Elektronikus könyvtári értelmező szótár. 1998. <<http://www.lib.uni-miskolc.hu/publ/ekszotar/>>
- [5] Dublin Core Metadata Element Set. Version 1.1. Reference Description. Recommendation. 1999. <<http://purl.org/dc/about/element-set.htm>>
- [6] GIBSON, W.: Neuromancer. Neuromanc. Budapest, Valhalla Páholy, 1992. 344 p.
- [7] GOLDEN D.–TÓTH T.–TURI L.: Virtuális örökkévalóság: objektumok a digitális könyvtárban. = Tudományos és Műszaki Tájékoztatás, 41. köt. 8–9. sz. 1998. p. 299–314. <<http://www.neuman-ház.hu/digitális/studies/object/objects.html>>
- [8] GÖZ Á.: Az Interneten elérhető információforrások katalogizálása. = Tudományos és Műszaki Tájékoztatás, 41. köt. 8–9. sz. 1998. p. 315–330. <<http://www.neuman-ház.hu/digitális/studies/intercat/index.htm>>
- [9] H'OBBSZAKON, R.: Internet evangelist: Hobbers Internet timeline, v1.1. 1994. 03. 12. <<http://infor.isoc.org/quest/zakon/Internet/History/Hit.html>>
- [10] HORVÁTH T.: A második cranfieldi jelentés. = Babiczky B.: Szövegyűjtemény az osztályozás és indexelés kérdéseinek tanulmányozására. Budapest, Tankönyvkiadó, 1970. p. 213–230.
- [11] Hypertext + multimédia. Oktatási segédanyag. Vál., az elősz. írta Sugár J., szerk. Klaniczay J. Budapest, Artpool, 1996. p. 29.
- [12] Das Internet und seine Dienste. <http://www.ion.de/service/tutorial/tutorial_3_2.htm>
- [13] KOCH, T.: Suchmaschinen im Internet. Vortrag auf der 1. INETBIB-Tagung, Dortmund, 1996. márc. 11. <<http://www.ub2.lu.se/tk/=demos/DO9603-manus.html>>, illetve angol változat <<http://www.ub2.lu.se/tk/=demos/DO9603-meng.html>>
- [14] KRISTULA, D.: The history of the Internet. 1997. <<http://www.davesite.com/webstation/net-history.shtml>>
- [15] LANDOW, G. P.: Hypertext and critical theory. New York, The Johns Hopkins University Press, 1992. <<http://twine.stg.brown.edu/projects/hypertext/landow/ht/contents.htm>> Magyarul az egyik fejezete jelent meg: Hypertextuális Derrida, posztstrukturalista Nelson? = Klaniczay J.–Sugár J.: Hypertext + multimédia. Oktatási segédanyag. Budapest, Artpool, 1996. p. 23–46.
- [16] LEINER, B. M. [et al.]: A brief history of the Internet. <<http://www.isoc.org/Internet/history/brief.html>>
- [17] MAGYAR K.: Tárgyszavazási gyakorlat és a tárgyszavas keresetőség az Interneten elérhető könyvtári állományokban. Budapest, Neumann-ház, 1999. <<http://www.neuman-ház.hu/digitális/studies/targyszavak/index.html>>
- [18] NOTESS, G. R.: Search engine showdown. The user's guide to the web searching Search engine statistics. = Notes, G.: A Greg Notes web site. 1999. <<http://www.notess.com/search>>

- [19] Az OMIKK Virtuális Könyvtára. Szerk. Válas Gy., Horváth P. 1999. 08. 16. <<http://www.omikk.hu/omikk/virkonyv/inet.htm>>
- [20] STEFFENS, U.: Information Retrieval: Grundlagen, Systeme und Integration. = Bibliothek, Forschung und Praxis, 21. köt. 1. sz. 1997. p. 54–58.
- [21] Suchen im Internet. <<http://www.privat.schlund.de/B/BesserSuchen/t-suchen.htm>>
- [22] SZAKADÁT I.: Xanadu. A Xanadu és Ted Nelson. <<http://www.uniworld.hu/netskills/tudas/HTML/Xanadu.htm>>
- [23] Xanadu. Project Xanadu. Founded 1960. The original hypertext project. 1999. 02. 02. <<http://www.xanadu.net>>

Jegyzetek

- ¹ Az internetre vonatkozó fontosabb testületi forrásokat, kronológiát stb. lásd [19].
 - ² Az Art Pool által 1996-ban kiadott hézagpótló szövegyűjteményben [11] Vannevar Bush, George P. Landow, Theodor Holm Nelson, és a hipertexttel foglalkozó számos más szerző tanulmányának fordítása található. Ugyanitt a hálón található történeti adatbázisok alapján magyar nyelvű kronológia, továbbá forrásjegyzék is megjelent. További, az internet történetét tartalmazó források [9], [14] és [16].
 - ³ Magyarul [11].
 - ⁴ [11], p. 56.
 - ⁵ Nelson életútját és a Xanadu történetét Szakadát István részletesen leírta [22].
 - ⁶ [11], p. 2.
 - ⁷ A rendszer önállósodva, szerényebb keretek között tovább él, több megvalósítása is megtalálható. Ezekről részletesebben a Xanadu honlapja és a róla elérhető Xanadu Australia cég honlapja tájékoztat [23].
 - ⁸ Az angol–magyar terminológiáról Drótos László állított össze a teljesség és véglegesség igénye nélkül szótárt [4].
 - ⁹ Barabasi, A.-L.–Albert, R.–Jong H.: Diameter of the World Wide Web. = Nature, 1999. 401. sz. szeptember 9. p. 130.
 - ¹⁰ Magyar példa a Hungary.Network egyszerű, jól áttekinthető gyűjtőszolgáltatásának a portálja <<http://www.net.hu/search>>, és az AltaVizsla indexelőszolgáltatását is magába foglaló Origo összetett felépítésű portálja.
 - ¹¹ Egy másik gyűjtőszolgáltatás, pl. a CUSI csoportosítva és magyarázattal is ellátva adja meg a különféle típusú keresőszolgáltatásokat, de még szótárakat is, és belőle kiindulva bármelyikben keresni lehet.
- Az ELTE TFK könyvtár–informatikai tanszékén is összeállítottak linkgyűjteményt. Szerzőik bíznak a nyelv rugalmasságában: az indexelőszolgáltatásokat az „index” osztályba, a katalógusokat a „mutató” osztályba sorolták. Ez olyan, mintha a kutyákat a „Canis”, a velük „ellentétes” macskákat pedig a „kutyák” osztályba sorolnák.

- ¹² Az interneten elérhető magyarországi könyvtárak tárgyszavaival foglalkozik [17].
- ¹³ A metaadatokra vonatkozóan lásd [5]; a digitális objektumokkal részletesebben foglalkozik [7] és [8].
- ¹⁴ Nem tévesztendő össze magával a Yahoo! internet-katalógussal. A Yahoo! Search e katalógus állományát indexelő rendszer, vele a katalógusban nem szisztematikusan (azaz nem böngészve), hanem egyedi szavakkal lekérdezve lehet keresni.
- ¹⁵ 1998-as adat [21].
- ¹⁶ Harvest Indexer, release 1.5. Hírcsoportja a következő címen található: <comp.infosystem.harvest>. A rendszerre vonatkozó publikációkat az alábbi címen archiválják: <http://www.mathematik.uni-osnabrueck.de/harvest/brokers/CIH>
- ¹⁷ Az interneten elérhető magyarországi könyvtári adatbázisok természetes nyelven alapuló keresettségével részletesen foglalkozik [17].
- ¹⁸ Ha internetkatalógusok (pl. Yahoo!) indexelőszolgáltatásairól van szó, a „Search” kiegészítővel szerepelnek (pl. Yahoo! Search).

Idézett keresőszolgálatok¹⁸

ALIWEB. Advertising Technologies Co.
<http://www.aliweb.com>

AltaVista. AltaVista Inc. <http://www.altavista.com>

AltaVizsla. Budapest, Origo, 1998.
<http://www.altavizsla.origo.hu>

Chemsoc <http://www.chemsoc.org/help/help.htm>

CUSI (Configurable Unified Search Engine).
<http://www.newton.ax.ac.uk/general/cusi.html>

Cyber Dewey. A guide to Internet resources organized using Dewey Decimal Classification codes. 1989.
<http://ivory.lm.com/1mundie/DDHC/DDH.html>

DIALOG Web. <http://dialogweb.com>

EELS (Engineering Electronic Library).
<http://www.ub2.lu.se/eel/eelhome.html>

Linkgyűjtemény. Budapest, ELTE TFK könyvtár-informatikai tanszék.
<www.tfk.elte.hu/konyvtar_tanszek/link/keresok.html>

Excite. Search <http://www.excite.com>

Google! <http://www.Google.com>

Harvest Home Page Broker.
<http://harvest.transarc.com/broker>

Harvest Indexer, release 1.5. Internet Research Task Force Group on Resource Discovery (ITRFG-RD), 1999. <www.tardis.ed.ac.uk/harvest>

Heuréka. Budapest, Hungary.Network, 1996.
<http://www.net.hu/search>

HotBot (HotWired). <http://www.hotbot.com>

HUDIR. Budapest, Hungary.Network, 1996.
<http://www.net.hu/search>

Infosec. <http://www.inforseek.com>

InterCat/WorldCat (OCLC Online Union Catalogue).
<http://www.oclc.org/oclc/man/6928fsdb/worldcat.htm>

Karlsruher Virtueller Katalog. Karlsruhe, Universität Karlsruhe. <http://www.ubka.uni-karlsruhe.de/krk.translate.html>

Kolibri. <http://www.kolibri.de>

KözElKat. <http://www.bibl.u-szeged.hz/kozelkat>

Lycos. Search <http://a2z.lycos.com>

Magellan. Search <http://www.mckinley.com>

MetaCrawler. <http://www.webcrawler.com>

MetaGer. <http://meta.rzn.uni-hannover.de>

MSN Web Search. <http://search.msn.com>

Music Search (MUSE). <http://www.ub.uni-heidelberg.de/helios/epub/info/daba/mnse.html>

NetFirst. OCLC <http://www.oclc.org/oclc/netfirst>

Northern Light. <http://www.northernlight.com>

Open Text Web Index. <http://index.opentext.com>

Snap! <http://www.dnsp.com>

Suchdienste im Internet. = Search & find in the Internet.
<http://www.bsz-bw.de/links/suchdienste.html>

Webcrawler. <http://www.webcrawler.com>

WWW Women. <http://www.women.com>

WWW Worm. <http://www.cs.colorado.edu>

Yahoo! Yahoo! Inc. Search <http://www.yahoo.com>

Beérkezett: 1999. IX. 27-én.

Állást kínál

A Medea Services Kft. könyvtárszakon végzett, németül jól beszélő fiatal **asszisztenst** (hölgyet) keres. Svájci tulajdonú cégünk Budapesten nyugat-európai könyvtárak részére könyvtári katalógusok konverziójával foglalkozik, és éppen egy rendkívüli nagy, többéves német projekt lebonyolítását szervezi. Ha a számítógép használata felhasználói szinten nem jelent problémát, felelősségteljes munkát szeret végezni, és önálló feladatokat keres, jó szervezőképesség és nyitottság jellemzi, és egy fiatal csapatban szeretne dolgozni, szívesen várjuk önéletrajzát címünkre: 1025 Budapest, Szalamandra u. 16/b, **dr. Pintér Marc** ügyvezető igazgató. Tel.: 325-5115 Fax: 325-8015 E-mail: medeaserv@matavnet.hu

Makovecz Imre műhelye

A Mundus Magyar Egyetemi Kiadó a rendkívüli érdeklődésre tekintettel új, jelentősen bővített kiadásban jelenteti meg a Gerle János szerk.: **Makovecz Imre műhelye** c. életműalbumot. Előfizetőknek jelentős kedvezmények. További tudnivalók a kiadó székhelyén: 1083 Budapest, Ludovika tér 2.; tel.: 210-1075/1122, 1185, 1198; fax: 333-5378; e-mail: werpe9812@matavnet.hu.